

March 2006

Nebraska Department of Education
SAMPLE PORTFOLIO

**DISTRICT ASSESSMENT SYSTEM DESCRIPTION
FOR READING, LISTENING, & SPEAKING**

DISTRICT ASSESSMENT SYSTEM DESCRIPTION:

Please describe your District Assessment System in this box. Include only those circumstances that directly impact the development of your district's assessment. (Ex: Any unusual constraints or issues related to the development of your assessment.)

Nebraskaland Public schools is located in the central part of the state. The district consists of two buildings – one elementary building, one 7-12 building. The elementary staff consists of 15 staff members and one principal. Included in the 15 staff members are one counselor who also teaches Title I and one special education teacher. Everyone was involved in the assessment process and teachers worked closely with the local Educational Service Unit. The principal provided much leadership to the teacher teams.

Quality Criterion One:

Assessments reflect the state or local standards.

- Districts should provide evidence that assessment content (test items or tasks) corresponds to the appropriate standards. Evidence of the match of assessments to the standards is based on an independent review of the alignment of assessments to the standards.
****Please be sure to clearly designate the independence of the review panel.**
- Assessments must be shown to be sufficient or comprehensive enough so that the determination of appropriate student classification into performance levels may be done.

In meeting this criterion, districts should answer and document the following questions:

1. **Who** did the process?

Include the number of panelists, their years of experience (collective estimates are appropriate), the grade levels that are represented and in what configurations: grade level, grades above or below the assessed grade, whether the groups were K-12 in nature. Include the number of teachers who participated in comparison to numbers of total staff. This describes the adequacy of the representation.

If the above panel is the group that assembled or wrote the assessments, it is important to have a second, independent review of the match to standards. If you have more than one panel, state it and describe it. **You will want people with independent, objective judgment to review assessment items or activities for match to standards.**

Describe who led your process. Was it a teacher leader, an educational service unit staff developer, an administrator? Include qualifications.

2. **What** did they do in this process?

Describe how the panel matched assessment items or activities to the standards. For example: How and when was this match reviewed by an independent group? What steps were actually taken? Include copies of the forms that the group used.

Describe the process used by the panel to determine that sufficient coverage or score points was present. For example: How was it decided that the coverage was comprehensive enough or sufficient to classify students into multiple performance levels? Were

there a required number of items or activities per standard at different difficulty levels? Were other assessment opportunities examined? What process was used? Include any forms the group used.

If you surveyed staff or collected data from others, indicate how you did that, the process that you used and include any forms that you may have used.

Independent Review Process:

The work was scheduled on a district curriculum day. The assessments were examined by the independent review team for match to standards, for numbers of items /performances and for the range of difficulty of those items /performances. Mrs. Principal facilitated and led the process.

Step One: The team decided to first make decisions independently about each assessment's match to standards, and then come back together as a group to discuss the decisions. Each person recorded his or her position on each item or task for each assessment and after everyone was done with each assessment, the team then discussed the independent ratings as a group and came to consensus. The consensus judgments were combined on a separate review form (include copy of form). Changes were recommended to the development team and documented for any item/task that did not match the standard. A checklist recorded the final decisions of the independent team.

Step Two: The independent review team examined each assessment for match to standard. At the top of each assessment were the standards measured in the assessment. Individually, each member of the review team made a judgment about each item or task relative to the standard using the review form described above. If they believed that the content and cognitive level of the item or task matched the standard, they placed an 'X' in the corresponding column. If they believed the item or task did not match either the content or the cognitive level, they placed a 'O' in the corresponding column.

Step Three: The independent review team members also examined each assessment to determine whether there were sufficient opportunities for placing students into a range of performance levels. This was done to ensure that all students, regardless of ability, would have an opportunity to demonstrate knowledge of the standards. The first thing the team did was to revisit the performance level definitions developed before the assessments were created. After Mrs. Principal reviewed those definitions with the team, the team members were asked to make a judgment about each item or task on each assessment a make a judgment about the appropriate performance level. For objectively-scored assessments, the team made note of the actual number of items that were at the beginning level, the progressing level, the proficient level, and the advanced level, indicating any missing or over-represented levels. For subjectively-scored assessments, the team examined the characteristics defined in the scoring rubric and the criteria for performance, making a determination of whether or not unique opportunities were found within the assessment task for students to perform at all four levels. After each team member made an independent decision, the team discussed their findings and made recommendations to the development team for changes. The following summary chart that follows documents the changes that were made following this review.

Step Four: Each member of the independent review team also examined the items and tasks for each assessment to determine whether there were enough items or tasks on the assessments used so that a student's knowledge or skill on the standard could be measured. In this case, the actual number of items or performances was recorded for each assessment. The review panel knew that scores would be used to classify students in four performance levels (beginning, progressing, proficient, and advanced), so they chose to require at least six items or tasks to measure each level of performance. In some instances multiple assessments were used to measure a standard, thus, the total number of items or tasks that were used to make a decision about student performance was included in the count. A summary chart that displayed the number of items or tasks that were judged to measure each level of performance was completed, and any notations of needed additions or deletions were made.

Step Five: The results of the independent review and recommended changes were shared with the assessment development team. Both the assessment development team and the independent review team had a chance to discuss the needed changes and agreement was reached as to who would make those changes. The summary chart that follows documents the results of the item or task match to the standards and evidence of the number of items or tasks that measure each performance level for each assessment.

3. ***What*** were the **results** of the process and **how** did you act upon them?

Describe what the results of the alignment process were. What happened as a result of the committees working together? What number of items or types of assessments was changed? What was kept? What was thrown out? What was learned? What was decided? Why? This is an essential step.

The summary chart below contains the results of the process including the recommended changes. All of the changes including additions, deletions, and revisions suggested by the independent review panel were made by the assessment development team.

*Criterion One***ASSESSMENTS MUST MATCH STANDARDS**

	ASSESSMENT	TYPE		NUMBER	OPPORTUNITY TO DEMONSTRATE				CHANGES MADE
Targets to be measured	Name	Obj.	Subj.	No. of Items/ Tasks	Beg.	Prog.	Prof.	Adv.	
4.1 Reading Strand (all reading standards)	Our Town, Short Story, and Social Studies Assessments	X		28 (obj)	10	7	8	3	Need to add more opportunities for students to demonstrate performance at the advanced level.
4.3 Speaking Strand (all speaking standards)	Native American Presentation		X	1 (subj.) task with a 4 pt. rubric	1	1	1	1	Need to add more opportunities for students to demonstrate performance at all levels.
4.4 Listening Strand (all listening standards)	Native American Presentation		X	1 (subj.) task with five 4 pt. rubrics	2	3	9	6	Need to add more opportunities for students to demonstrate performance at the beginning and progressing levels.

SAMPLE

Quality Criterion Two –

Students have an opportunity to learn the content.

- Districts should provide evidence that assessment content (test items or tasks) corresponds to the district’s curriculum and instruction. The match of assessments to instruction is based on evidence that the content of the assessments has been taught.
- Instruction must have occurred prior to assessment. For this criterion to be fully met, students should have instruction on 80% or more of the content prior to being assessed on that content.
- **“Opportunity to learn” as described in this criterion means that the local school district’s curriculum and instruction offers the opportunity for students to meet the standards prior to the assessment. It means the district-collected evidence showing students had the opportunity to learn. Documentation of that evidence is provided in this criterion.**

*Note: Even if you are using the norm referenced tests aligned to the Nebraska standards by the Buros Center for Testing, you **still** need to describe how Criterion 2 was met.*

In meeting this criterion, districts should answer and document the following questions:

1. **Who** did the process?

Include the number of panelists, their years of experience (collective estimates are appropriate), the grade levels that are represented and in what configurations: grade level, grades above or below the assessed grade, whether or not the groups were K-12. Include the number of teachers who participated in comparison to numbers of total staff.

If you compiled information from teachers or others and used forms or questionnaires to collect that data, include a copy of all forms that were used to include information from others.

Describe who led your process. Was it a teacher leader, an educational service unit staff developer, an administrator? Include qualifications of the leader.

Two of the teachers from the Assessment Development team, Mrs. Reviewer Three and Mrs. Reviewer Four along with Mrs. Principal (See Criterion One for qualifications) attended the sessions at the Educational Service Unit's STARS Academy. They learned the process of curriculum alignment with the standards, and then returned to the district and shared the process with all fifteen of the elementary staff. At the ESU STARS Academy the curriculum alignment process was led by Mrs. Staff, a former elementary teacher, and Mrs. Developer.

During a district in-service day all fifteen elementary staff were required to examine the local curriculum for the standards and participated in the process outlined below. The elementary staff worked separately from the secondary staff.

2. **What** did they do in this process?

Describe how your panel(s) examined your local curriculum to find where standards were covered. For example: If you were identifying reading, speaking, and listening standards in another curriculum area, indicate that. Indicate the steps that your teacher teams took.

Timing is an important component of this criterion so it must be demonstrated that assessments are given after sufficient instruction has occurred.

If you examined textbooks and instructional materials, describe how you determined when during the school year, the standards were covered. If you collected information through lesson plans, collected classroom assignments, classroom assessments, or classroom observations, give specific examples about what percentage of teachers were involved, who examined these products or made the observations, and how it was determined that the content of assessment was being taught. If you used surveys or questionnaires, include the forms used for that process.

Describe how you determined the point in time when assessment(s) would occur. Was it throughout the year? Was it at one point in time after enough instruction had occurred? ****Dates or a clear indication of timing is important and required.** Indicate at what point in the year the instruction and the reading assessment occurred – give the month(s) of both teaching and assessment administration and describe why you made that decision. Provide the rationale for the decision and the process used to determine that rationale.

Step One: Each staff member had a copy of the reading, speaking, and listening standards and a copy of the most current curriculum guide. The staff members met in small groups with their grade level neighbors: The kindergarten teachers, the first grade teachers and the counselor worked together; the second grade and third grade teachers worked with the Title I teacher; the fourth grade and fifth grade teachers worked with the special education teacher.

Mrs. Principal had prepared an alignment worksheet. (include copy of the form) Using the worksheet as a guide, each group of teachers identified in what unit each standard was taught, when it was taught and when it was assessed. As they identified the standards within the curriculum, they made note of any curriculum holes, gaps, or overlaps that were present. Sample completed worksheets are attached.

Step Two: After each small group had worked together to complete the worksheets for the alignment process, the larger group came back together and shared the curriculum holes, gaps, and overlaps. They discussed the changes needed in the local curriculum. The group decided how those changes would be addressed and who would make those needed adjustments.

Step Three: Lastly the group discussed how they would ensure the opportunity for students to learn. They wanted a method that would not be too time consuming or cumbersome. They thought the simplest method might be to document the teaching of the standards in their lesson plans. Mrs. Principal suggested some minor modifications and coding in their lesson plan books that would make recording the teaching of standards relatively simple. The teachers decided to try that new coding for the upcoming year and see how it worked out.

Mrs. Principal indicated she would be dropping into classrooms to do regular classroom observations and walk-throughs and would expect to see the lesson plans when doing her visits. She indicated also that she would be randomly collecting lesson plans, that they should be available at all times. Included in this portfolio is a sample lesson plan, and the schedule of observations made by Mrs. Principal (include appropriate documentation.)

3. What were the results of the process and how did you act upon them?

Describe the results of the curriculum alignment process. What happened as a result of the committees working together? Or the surveys? Or the collection of lesson plans? What was changed? How was instruction or classes offered changed or adjusted? What was learned? What was decided? Why? Identify and describe the outcomes. Were there gaps or overlaps? State what you found; give the results of your work. Indicate what you intend to do with your findings.

The chart below is a completed summary of the alignment and opportunity to learn process, including the suggested changes that were made by the elementary staff.

*Criterion Two***OPPORTUNITY TO LEARN**

Standard	Unit Taught (When)	Assessment Used	When Assessed	Changes Made
4.1 Reading Strand (all reading standards)	All fall, specific units in January -February	Our Town, Short Story, and Social Studies Assessments	March 8-10	Moved section of curriculum from fifth grade to fourth grade to provide opportunities to learn before assessment.
4.3 Speaking Strand (all speaking standards)	November, March-April	Native American Presentation	April 22-26	
4.4 Listening Strand (all listening standards)	December, March-April	Native American Presentation	April 22-26	Formalized the instruction of listening.

SAMPLE

Quality Criterion Three

Assessments are free from bias or offensive language or situations.

- Districts should provide evidence that assessment content has been examined critically to be free of bias against any group of students. Assessments should have been examined to be free of offensive language and stereotyping. For this criterion to be met, there needs to be a staff development component that includes an orientation to bias for the assessment developers and an examination of the content of the assessment(s) by an independent panel that has been trained to conduct a bias review or a statistical analysis that examines the differences in performance of selected groups of examinees or both.

In meeting this criterion, districts should answer and document the following questions:

1. **Who** did the process?

This includes a staff development component so that those writing assessment tasks/items and independent reviewers are aware of biased or sensitive situations. Describe who led the staff development and include the qualifications of the leader(s).

If your district used a panel of experts from outside the school district, include the number of panelists, their qualifications and why it was important to have them on the panel. If you used teacher panels, summarize the characteristics of the panel including their years of experience, the grade levels that are represented and in what configurations. If panel members were included because of diversity, indicate what that was and why.

Include any forms that the independent review panel used to review assessment items or tasks for bias and sensitive language.

An elementary team attended the bias training at the Educational Service Unit. The training was led by Dr. Expert, whose qualifications include a Ph.D. in measurement and over ten years experience with educational assessment. The elementary team members were as follows:

Bias Reviewers:

Mrs. Reviewer Four	Grades 3-6	11 yrs. experience
Mrs. Reviewer Five	Grades 4-5	5 yrs. experience
Mrs. Reviewer Six	Grades K-6	15 yrs. experience
Mrs. Principal	Administrator	9 yrs. experience

2. **What** did they do in this process?

Describe the process used to insure that items/activities in the assessment(s) are free of bias. Describe the bias “orientation” process as well as the “assessment examination” process. It is not appropriate to merely “accept” the assurance of a textbook company or of a test publisher that items are bias-free or not sensitive¹. Items/activities need to be examined locally through your process to eliminate or change any items of an inappropriate or sensitive nature. How did the teams judge whether items or tasks were free from bias?

If you used a statistical method to assess bias, specify and describe the method used.

Step One: The team attended a bias training day at the service unit. Dr. Expert provided a PowerPoint™ overview about fairness in testing. Basically, the team learned about fairness in the categories of sensitive situations and studied examples of varying kinds of bias and unfairness. The team joined teachers from other districts to examine sample assessments and to make determinations about whether or not items or tasks might be unfair in any way whether in language, situation, or unfair to any group of people.

Step Two: Our district team exchanged assessments with a team from another district. A worksheet, developed by the Service Unit was completed as each assessment was reviewed for bias. Any recommended changes were made and recorded on the sheets. Sample completed worksheets are included.

Step Three: At the close of the day, each district team received the completed worksheets about their own assessments and had the opportunity to discuss the recommended changes.

3. **What** were the results of the process and **how** did you act upon them?

Describe the results of the bias review. What was changed? What was thrown out? Your evidence is strengthened by giving examples of items or tasks that were revised or updated.

The summary chart including the recommended changes were reviewed and acted upon by the assessment development team.

¹ Note that if a district is using a Norm-Referenced Tests included in NDE’s alignment studies, it does not need to include information about a local bias review.

Criterion Three

ASSESSMENT HAS BEEN REVIEWED FOR BIAS Training Date(s): June 15, 2002

ASSESSMENTS	DATES	CHANGES MADE
	REVIEW	
4.1 Reading Strand (all reading standards)		
Our Town Assessment	June 15, 2002	Selected different jobs to be examined so they weren't gender biased – both male and female executives.
Short Story Unit Assessment	June 15, 2002	Clarified directions – too vague, no changes on any items.
Social Studies Assessment	June 15, 2002	Changed an item that may have been offensive to Roman Catholics, changed two items that may have been offensive to people from the Middle East.
4.3 Speaking Strand (all speaking standards)		
Native American Presentation	June 15, 2002	Revised rubric so that students' appearance was not considered.
4.4 Listening Strand (all listening standards)		
Native American Presentation	June 15, 2002	References to "Indian" in two of the rubrics were changed to Native American.

SAMPLE

Quality Criterion Four –

The level is appropriate for students.

- Districts should provide evidence that assessment content has been examined critically to determine the appropriateness of the level. This criterion can be met if there is evidence that a panel of individuals qualified to determine the level of appropriateness, examined the content of the assessment(s) and found it to be appropriate. The criterion may also mean an appropriate reading level for the standard being assessed (for assessment tasks that require students to read).

In meeting this criterion, districts should answer and document the following questions:

1. **Who** did the process?

Describe the panelists, their years of experience, the grade levels that are represented and in what configurations: all the same grade level, grades above or below the assessed grade, anything that would show that these panelists would be qualified judges of appropriate level of tasks for students. Include the number of teachers who participated in comparison to numbers of total staff. Indicate the point in time when the groups met.

For example: If the school psychologist or others in your district who have expertise in child development participated, indicate their involvement and/or who led the process. Include the qualifications.

If someone conducted a readability analysis, indicate which analysis method was used and why it is appropriate.

(Note that readability analyses are not appropriate for speaking and listening assessments and may not be really appropriate for math, science, or social studies assessments because the reading requirements are likely too limited to estimate accurately the readability level.)

On a designated staff development day, the independent review team (see Criteria One for qualifications) met to review the assessments for appropriateness of level. The discussion was led by Mrs. Principal whose qualifications are also described in Criterion One.

2. **What** did they do in this process?

Describe the process you used to review items to make sure they were at an appropriate level. What steps did you take? How did you make this determination? For example: Describe how you collected the evidence. If you used questionnaires or forms to collect data, include copies of those forms.

If you used a statistical analysis of readability level, indicate which one (s) was used and why the process was selected.

Note: Relying on a textbook's statement about readability without supporting evidence is not adequate.

Step One: The team reviewed all of the assessments to make a determination about the appropriateness of the assessments. Mrs. Principal asked that each team member make an individual decision about the appropriateness of the assessment task level of the vocabulary and of the materials used. Any recommendations for changes to improve appropriate level were recorded by the individual team members. (Examples of the worksheets that the independent review used to make these judgments should be included.)

Step Two: After each team member independently reviewed each assessment using the criteria described above, the independent review team discussed the results together coming to a consensus about whether or not the items or tasks for each assessment were appropriate and recommended changes to be made in the assessments to the assessment development team. The summary chart provides information about these changes and how the assessment development team responded to them.

Step Three: The Fry readability analysis was used on the materials as well. The results are reported on the next page. The results of the social studies readability caused the team to select new passages.

3. **What** were the results of the process and **how** did you act upon them

Describe what discoveries you made in the process and any modifications that you determined were needed on the basis of recommendations of the panels or qualified reviewers. Indicate any changes that you made as a result of the review. If your district conducted a readability analysis, include a description of the results and the changes that were made.

The chart below indicates the results of the review for appropriate level. It should be noted that all changes recommended by the reviewers were made.

*Criterion Four***THE LEVEL IS APPROPRIATE FOR STUDENTS**

ASSESSMENTS	Readability	Panel Cognitive Level Review	CHANGES MADE
4.1 Reading Strand (all reading standards)			
Our Town Assessment	4.3 (Fry)	X	No changes were made to items or reading passages.
Short Story Unit Assessment	4.4 (Fry)	X	Two questions were revised to measure higher ability skills
Social Studies Assessment	8.6 (Fry)	X	No changes were made to the items, but the panel judged the reading level of the passages too high. New passages will be selected to reflect more grade appropriate levels. The new passages will be reviewed by the panel.
4.3 Speaking Strand (all speaking standards)			
Native American Presentation	N/A	X	No changes were made to the directions, assessment task, or the rubric that were examined in the cognitive level review.
4.4 Listening Strand (all listening standards)			
Native American Presentation	N/A	X	We changed one of the rubrics to better articulate the skills that would be demonstrated by beginning and progressing level students.

Quality Criterion Five

There is consistency of scoring.

Districts should provide evidence that assessments are reliable in terms of the consistency of scores that might be expected of students who might take the assessment(s) on more than one occasion without intervening instruction. It also addresses the consistency of scoring for subjectively scored items, prompts, or performance.

- Districts should provide evidence of reliability for assessments that include multiple choice or other objectively scored items. The level that will result in this criterion being rated as fully met, is a consistency measure of .70 or higher (averaged across all assessments).
- If you have used an appropriate process for calculating reliability but have an average reliability value lower than .70 and provide a feasible plan that is likely to improve reliability, you will receive a “Needs Improvement.”
- If you have used an appropriate process for calculating reliability, but the average reliability is lower than .70 and your improvement plan is not likely to improve reliability or you do not provide any plan to improve the reliability, you will receive a “Not Met” rating. If you have not used an appropriate process to calculate reliability, you will receive a “Not Met” rating.
- Evidence of reliability should also be provided for subjectively scored assessments. For these types of assessments (e.g., essays, student research papers) there should be clear scoring criteria (a scoring guide or rubric), evidence should be provided that there was training on the scoring rubric, and there should be evidence that the scoring criteria are applied consistently. This is typically demonstrated by showing that multiple scorers have independently agreed with scoring a sample of student work.

Districts may receive one of four ratings:

“Met”

“Needs Improvement”

“Met some further comment necessary”

“Not Met”

- A “Needs Improvement”, or “Met-some further comment necessary” will be considered the same as “Met” except in the case of Criterion #5.
 - If Criterion 5 receives a “Met-some further comment necessary” it will be considered the same as “Met,” and it is possible to receive an exemplary rating.
 - If Criterion 5 receives a “Needs Improvement,” it is not considered the same as “Met”, and the highest rating possible will be “Very Good.”

In meeting this criterion, districts should answer and document the following:

1. **Who** did the process?

Indicate the person(s) or teams who actually led, conducted, and participated in the reliability analysis. Include their qualifications.

Mrs. Principal, the elementary administrator with 9 years of experience and training through the service unit, assumed the responsibility for calculating the reliability estimates. She attended the Educational Service Unit's workshop on Criteria #5 and #6 and had obtained the necessary knowledge and skills to assume this responsibility.

2. **What** did they do in this process?

There are several ways to measure the reliability of scores on a district's assessments. Some ways are more appropriate for some assessment methods and size of student population than others. The methods described below are only suggestions. Other methods not listed include: alternate forms, internal consistency such as coefficient alpha, KR20, KR21, and split-half, and test-retest.

Reliability for all assessments was calculated by assessment using a decision consistency model (for objectively scored assessments) and inter-rater reliability (for subjectively scored assessments).

Decision Consistency (Short Story Assessment, the Our Town Assessment, and the Social Studies Assessment)

Step One: Mastery levels (See Criterion Six) were established using definitions of mastery levels that were previously agreed upon. Through the process described in Criterion Six, mastery levels for beginning, progressing, proficient, and advanced performance levels were set.

Step Two: The teachers made a professional judgment about each student's abilities on the content measured on each of the assessments, predicting whether students were at the beginning, progressing, proficient, or advanced levels of performance, remembering that only the definitions of performance were considered in these judgments. These judgments occurred after students had been instructed in the content measured by the assessment. After the assessment was administered students were classified into performance levels using the cut scores described in Step One. The extent that the teachers' predictions about performance and students' actual performance agreed exactly was calculated (include copy of the form used.)

Step Three: On any assessment with agreement below .70, Mrs. Principal gathered the teachers together for a discussion. The teachers talked about the differences between the expectations for students and the resulting outcomes. Specifically for the reading assessment, most misclassifications of students' performance were between the proficient and advanced levels. Because

there were limited opportunities for the students to demonstrate advanced skills, the team decided to add more advanced items and give the advanced students an opportunity to be challenged. The team also revisited the performances level definitions of beginning, progressing, proficient, and advanced. In the discussion they discovered some differences in the teacher's thinking about the definitions of performance and other factors (e.g., student effort or attitude) that they might use in their grading practices. The team agreed that these external factors should not influence their judgments of performance in the future.

Inter-rater Reliability (Native American Presentation and the speaking portion of Our Town)

Step One: The Native American Presentation that had sections that were used to measure speaking and listening performance was double scored. In the case of these performances all elementary teachers were trained on the scoring rubrics for speaking and listening. Mrs. Principal led the training.

First, the team went through the rubrics describing the characteristics of performance for each scorepoint. Then she showed the team videotaped examples of various performances. Two team members independently scored each performance in the double scoring. All students' performances were included.

Step Two: During class the classroom teacher scored the performance as it occurred. Because the performances were videotaped, the second rater was able to score the performances at a later date. In this way all of the students were double rated.

Step Three: In all cases exact agreement of the two raters was calculated because the scoring rubric is only four points.

3. **Results** of the process and **how** did you act upon them?

Describe the reliability value that the reliability analysis produced for the assessments that you used. **If the reliability values are not as high as you would like them to be, describe what you will do to increase the reliability.** Your description should include examples of items that you need to change and specific steps you intend to take to change them. If your process is acceptable but the plan for improving reliability is not clearly described, the highest possible rating on Criterion 5 is "Needs Improvement." A "Needs Improvement" on Criterion #5 prevents an "Exemplary" overall rating on the portfolio. Note that if inappropriate processes were used or if the plan for improving reliability is not likely to be effective, this will result in a "Not Met" rating.

Criterion Five

THE ASSESSMENTS ARE SCORED RELIABLY

ASSESSMENTS	METHOD USED	RESULTS	CHANGES TO IMPROVE RELIABILITY
4.1 Reading Strand (all reading standards)	Decision Consistency (4 levels)	.72	We plan to add some more advanced level items to help us better distinguish between advanced and proficient performance levels. Most of the disagreements between our teachers' predictions and the students' performance occurred between the proficient and advanced levels.
4.3 Speaking Strand (all speaking standards)	Inter-rater agreement (4 pt. rubric)	.65 (exact)	The rubric had some inconsistencies in the progressing and proficient levels that duplicated information, making it difficult for the raters to distinguish between the levels. By eliminating this apparent problem with the rubric, and by highlighting this revision in the training activities, we expect that it will improve the level of agreement.
4.4 Listening Strand (all listening standards)	Inter-rater agreement (on total score – 20 points)	.46 (exact) / .77 (adjacent)	Although the exact level of agreement was lower than desired, the size of the scale (20 points) gave us confidence that scores of raters could be within one point of each other. Adjacent agreement was acceptable and no systematic patterns of disagreements surfaced after examining these. Thus, no changes are planned for the rubric design, rater training or scoring activities
	Average Reliability	.71 (mean) / .72 (median)	

Quality Criteria Six –

The mastery levels are appropriate.

Districts should provide evidence that they have used an appropriate process for determining “how good is good enough” in terms of levels of student achievement.

- Districts should provide evidence that the student mastery decisions were made using procedures that take into account the difficulty of the items or tasks in the assessments or classifications of students on an independent criterion (other than the assessment). The procedure used to set mastery levels should include systematic judgments about assessment content and the different levels of student performance.
- The important thing here is for districts to identify and describe the method used to set mastery levels. Districts may not rely on their traditional grading scale to make these decisions. Professional judgment about students or about the test/work itself need to be used to arrive at mastery level decisions.

Quality Criteria 1-4 focus on the assessment development phase. Quality Criteria 5 and 6 typically occur after assessments have been administered, and you have actual student results.

In order to meet this criterion, districts should answer and document the following questions:

1. **Who** did the process?

Include the number of panelists, their years of experience (collective estimates are appropriate), the grade levels that are represented and in what configurations: grade level, grades above or below the assessed grade, whether or not the groups involved teachers in multiple grade levels. Include the number of teachers who participated in comparison to numbers of total staff.

If you compiled information from teachers or others and used forms or questionnaires to collect that data, include a copy of all forms that were used to include information from others.

Describe who led your process. Was it a teacher leader, an educational service unit staff developer, an administrator? Include qualifications.

Mastery levels were determined by an elementary committee who have the students this year. Mrs. Principal facilitated the discussion. She had obtained her training through the Educational Service Unit. The qualifications are as follows:

Mrs. Reviewer Four	Grades 3-6	11 yrs. experience
Mrs. Reviewer Five	Grades 4-5	5 yrs. experience
Mrs. Reviewer Six	Grades K-6	15 yrs. experience
Mrs. Principal	Administration	9 yrs. experience

2. **What** did they do in this process?

Your district may have used one of the following processes or others that you have chosen. For example:

Modified Angoff Method for all reading standards and all listening standards.

Step One Mrs. Principal led the discussion. The team started with the performance level definitions that were used and previously described in Criteria 1 and 5. For this process, we focused on the characteristics of the “Barely Progressing,” “Barely Proficient,” and “Barely Advanced” students and what characteristics distinguished them from the performance level just below. For example, we talked about what distinguished the “Barely Progressing” student from the Beginning student for the content represented in each assessment, and so on. We wrote the characteristics for the objective assessments for each of these “target students” on chart paper and put them up around the room for the teachers to reference during the judgment process.

Step Two: Because of our small number of students, we used the modified Angoff Method which requires panelists to make judgments about how the “target students” will likely perform on each item.

Step Three: Next, Mrs. Principal had the team members examine each item on the assessments and make an independent determination (based on the definitions we had agreed upon) whether the barely progressing student will answer correctly, whether the barely proficient student will answer correctly, or whether the barely advanced student will answer correctly. The passing score for each level was determined to be the number of items that will be answered correctly by students at each level.

Step Four: The results were averaged across all the team members. We used the mean score for each level as our cut score (include a copy of forms used.)

For the speaking performances: - Analytical Judgment

Step One: Mrs. Principal had samples of video taped speaking performances. These sample illustrated what speaking performances would look like at all four levels.

Step Two: The team looked at 25 video taped performances and each team member classified them as one of four levels. Then from those classifications, each team member identified what they believed were the two best “beginning” performances and the two worst “progressing” performances. The actual scores from these performances were averaged to determine the progressing cut score. This process was reported for both the proficient and advanced cut scores.

For the listening performances – Analytical Judgment

Step One: Mrs. Principal provided samples of the listening responses that had already been scored but the team did not know the scores.

Step Two: The team looked at 25 of these performances and each team member classified them as one of four levels. Then from those classifications, each team member identified the two best “beginning” performances and the two worst “progressing” performances. As with speaking, the actual scores from these performances were averaged to determine the progressing cut score. The process was repeated for both the proficient and advanced cut scores.

Step Three: As in the case of speaking, the actual scores were averaged and the averages averaged across the team members results.

3. *What* were the **results** of the process and **how** did you act upon them?

Describe the results or any new learning in the process you used. Provide the actual numbers that will be used as your mastery levels.

Criterion Six

MASTERY LEVELS APPROPRIATE

ASSESSMENT	METHOD	MASTERY LEVELS				CHANGES MADE
		Beg.	Prog.	Prof.	Adv.	
4.1 Reading Strand (all reading standards)	Modified Angoff	0-17	18-21	22-26	27-28	
4.3 Speaking Strand (all speaking standards)	Analytical Judgment	1	2	3	4	We plan to make changes to the number of opportunities that students have to demonstrate performance at all levels.
4.4 Listening Strand (all listening standards)	Analytical Judgment	0-10	11-13	14-16	17-20	

SAMPLE

CRITERION 1: Alignment of Assessments to Standards*Who did the process?**Describe process...**Results and actions taken (complete table below):*

Standards	Type of assessment (CRT, NRT, Teacher Made)	Beginning	Progressing	Proficient	Advanced	Actions taken	Comments
4.1.1 Reading unfamiliar words and phrases							
4.1.2 Increase vocabulary							
4.1.3 Identify main idea and supporting detail							
4.1.4 Identify the appropriate resources							
4.1.5 Identify and use characteristics to classify types of text							
4.1.6 Knowledge of the structure, elements, and literary techniques to analyze fiction							
4.1.7 Knowledge of text structure and organizational elements to analyze nonfiction or informational text.							
4.1.8 Identify similarities and differences between two reading selections							
4.3.1 Participate in group discussions by asking questions and contributing information and ideas.							
4.3.2 Deliver organized oral presentations using complete sentences, clear enunciation, adequate volume, eye contact.							
4.4.1 Identify information gained and complete tasks through listening.							

Criterion 2: Opportunity to Learn*Who did the process?**Describe process...**Results and action taken (complete table below):*

Standard	Type of assessment, NRT, CRT, Teacher-made	State date of relevant instruction	End date of relevant instruction	Date of Assessment	Did assessment occur after instruction? (Y/N)	Were retests given? (Y/N)	Comments
4.1.1 Reading unfamiliar words and phrases							
4.1.2 Increase vocabulary							
4.1.3 Identify main idea and supporting detail							
4.1.4 Identify the appropriate resources							
4.1.5 Identify and use characteristics to classify types of text							
4.1.6 Knowledge of the structure, elements, and literary techniques to analyze fiction							
4.1.7 Knowledge of text structure and organizational elements to analyze nonfiction or informational text.							
4.1.8 Identify similarities and differences between two reading selections							

BEGINNING	PROGRESSING	PROFICIENT	ADVANCED
Demonstrates obvious difficulty reading and understanding text. Inconsistently uses strategies. Makes few if any connections between text and life experiences.	Demonstrates the use of limited strategies while reading, and may have a narrow understanding of text. Applies text to life experiences occasionally, but is sometimes unclear.	Demonstrates the ability to use multiple strategies to read and understand text. Makes clear connections between multiple selections and their own life experiences.	Demonstrates automatic command of concepts and integration of information. Makes complex inferences and interprets figurative language. Demonstrates the ability to generalize topics in the reading selection.

4.3.1	Participate in group discussions by asking questions and contributing information and ideas.							
4.3.2	Deliver organized oral presentations using complete sentences, clear enunciation, adequate volume, eye contact.							
4.4.1	Identify information gained and complete tasks through listening.							

SAMPLE

Criterion 2: Opportunity to Learn

Indicate with a line the time interval when each standard is taught and place a + (plus) at the time of assessment.

Standard	Aug.	Sept.	Oct.	Nov.	Dec.	Jan.	Feb.	March	April	May
4.1.1 Reading unfamiliar words and phrases										
4.1.2 Increase vocabulary										
4.1.3 Identify main idea and supporting detail										
4.1.4 Identify the appropriate resources										
4.1.5 Identify and use characteristics to classify types of text										
4.1.6 Knowledge of the structure, elements, and literary techniques to analyze fiction										
4.1.7 Knowledge of text structure and organizational elements to analyze nonfiction or informational text.										
4.1.8 Identify similarities and differences between two reading selections										
4.3.1 Participate in group discussions by asking questions and contributing information and ideas.										
4.3.2 Deliver organized oral presentations using complete sentences, clear enunciation, adequate volume, eye contact.										
4.4.1 Identify information gained and complete tasks through listening.										

CRITERION 3: Are the Assessments Free from Bias?*Who did the process?**Describe process...**Results and actions taken (complete table below):*

Standards	Type of assessment (CRT, NRT, Teacher Made)	Number of items on assessment	Number of items judged to be biased	Actions taken. (Describe what you did with all items identified as biased) For example: item tossed out (item discarded); item revised and reviewed again, then included in the assessment; no action taken.	Comments
4.1.1 Reading unfamiliar words and phrases					
4.1.2 Increase vocabulary					
4.1.3 Identify main idea and supporting detail					
4.1.4 Identify the appropriate resources					
4.1.5 Identify and use characteristics to classify types of text					
4.1.6 Knowledge of the structure, elements, and literary techniques to analyze fiction					
4.1.7 Knowledge of text structure and organizational elements to analyze nonfiction or informational text.					
4.1.8 Identify similarities and differences between two reading selections					
4.3.1 Participate in group discussions by asking questions and contributing information and ideas.					
4.3.2 Deliver organized oral presentations using complete sentences, clear enunciation, adequate volume, eye contact.					
4.4.1 Identify information gained and complete tasks through listening.					

Criterion 4: The Level is Appropriate for Students*Who did the process?**Describe process...**Results and action taken (complete table below):*

Standard	Type of assessment, NRT, CRT, Teacher-made	Number of items on assessment.	Number of items reviewed for appropriateness of level.	Number of items identified as appropriate for level.	Number of items identified as needing revision or	Actions taken. What did you do? For example: item tossed out (item discarded); item revised and reviewed again, then included in the assessment; no action taken.	Comments
4.1.1 Reading unfamiliar words and phrases							
4.1.2 Increase vocabulary							
4.1.3 Identify main idea and supporting detail							
4.1.4 Identify the appropriate resources							
4.1.5 Identify and use characteristics to classify types of text							
4.1.6 Knowledge of the structure, elements, and literary techniques to analyze fiction							
4.1.7 Knowledge of text structure and organizational elements to analyze nonfiction or informational text.							
4.1.8 Identify similarities and differences between two reading selections							
4.3.1 Participate in group discussions by asking questions and contributing information and ideas.							
4.3.2 Deliver organized oral presentations using complete sentences, clear enunciation, adequate volume, eye contact.							
4.4.1 Identify information gained and complete tasks through listening.							

CRITERION 5: There is Consistency in Scoring*Who did the process?**Describe process...**Results and actions taken (complete table below):*

Standards	Type of assessment (CRT, NRT, Teacher Made)	Number of items on assessment	Number of students	Reliability coefficient (needs to be at least 0.7)	Method Used KR21, KR20, Coefficient Alpha, a Decision Consistency method, etc.	Comments, indicate why you used the method you chose.
4.1.1 Reading unfamiliar words and phrases						
4.1.2 Increase vocabulary						
4.1.3 Identify main idea and supporting detail						
4.1.4 Identify the appropriate resources						
4.1.5 Identify and use characteristics to classify types of text						
4.1.6 Knowledge of the structure, elements, and literary techniques to analyze fiction						
4.1.7 Knowledge of text structure and organizational elements to analyze nonfiction or informational text.						
4.1.8 Identify similarities and differences between two reading selections						
4.3.1 Participate in group discussions by asking questions and contributing information and ideas.						
4.3.2 Deliver organized oral presentations using complete sentences, clear enunciation, adequate volume, eye contact.						
4.4.1 Identify information gained and complete tasks through listening.						

Angoff Method Worksheet

BEGINNING	PROGRESSING	PROFICIENT	ADVANCED
Demonstrates obvious difficulty reading and understanding text. Inconsistently uses strategies. Makes few if any connections between text and life experiences.	Demonstrates the use of limited strategies while reading, and may have a narrow understanding of text. Applies text to life experiences occasionally, but is sometimes unclear.	Demonstrates the ability to use multiple strategies to read and understand text. Makes clear connections between multiple selections and their own life experiences.	Demonstrates automatic command of concepts and integration of information. Makes complex inferences and interprets figurative language. Demonstrates the ability to generalize topics in the reading selection.

Criterion 6: The mastery levels are appropriate.

Standard	Teacher Judgment (Beginning, Progressing, Proficient, Advanced)	Actual Student Results	Changes Needed
4.1.1 Reading unfamiliar words and phrases			
4.1.2 Increase vocabulary			
4.1.3 Identify main idea and supporting detail			
4.1.4 Identify the appropriate resources			
4.1.5 Identify and use characteristics to classify types of text			
4.1.6 Knowledge of the structure, elements, and literary techniques to analyze fiction			
4.1.7 Knowledge of text structure and organizational elements to analyze nonfiction or informational text.			
4.1.8 Identify similarities and differences between two reading selections			
4.3.1 Participate in group discussions by asking questions and contributing information and ideas.			
4.3.2 Deliver organized oral presentations using complete sentences, clear enunciation, adequate volume, eye contact.			
4.4.1 Identify information gained and complete tasks through listening.			